MCAS project

Data integration workflow architecture.

A. Baranovski

Data integration workflow

- Workflow
  - Off line pre-processing, “artifact” generation(img. for ex)
  - Structure data transformations for re-usability
- Integration
  - Data com. protocol/format transforms(!)
  - Schema flexible storage
  - Query and aggregate
  - Validation

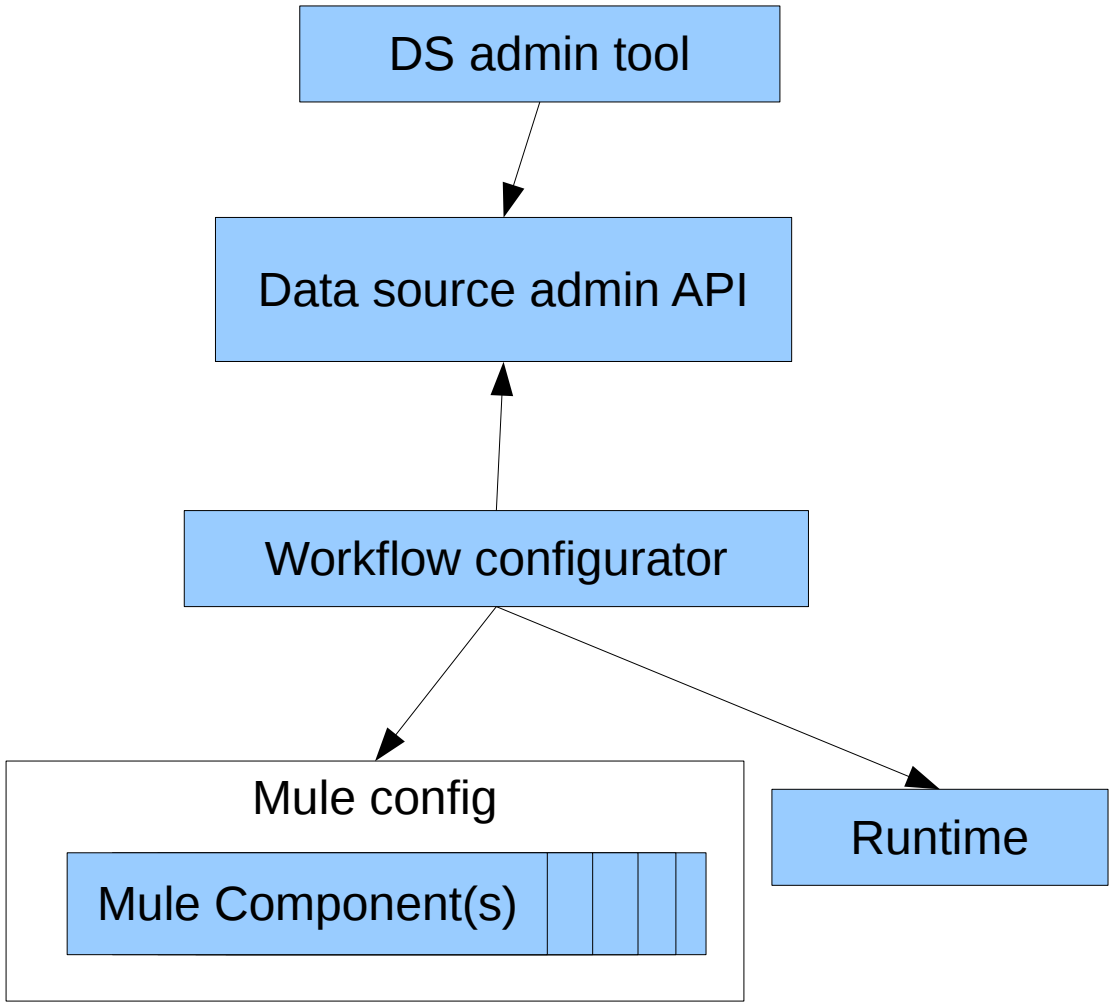
Candidate software

- Mule
  - MODEL and implement data and execution flow in a transport and component interface agnostic way.
    - Translates data formats
    - Auto discovers object interfaces and makes them available through:
      - Variety of communication protocols (HTTP is primary interest now)
    - SEDA
    - It works and is easy to configure and define concepts in.
- XML db
  - Schema flexible data storage

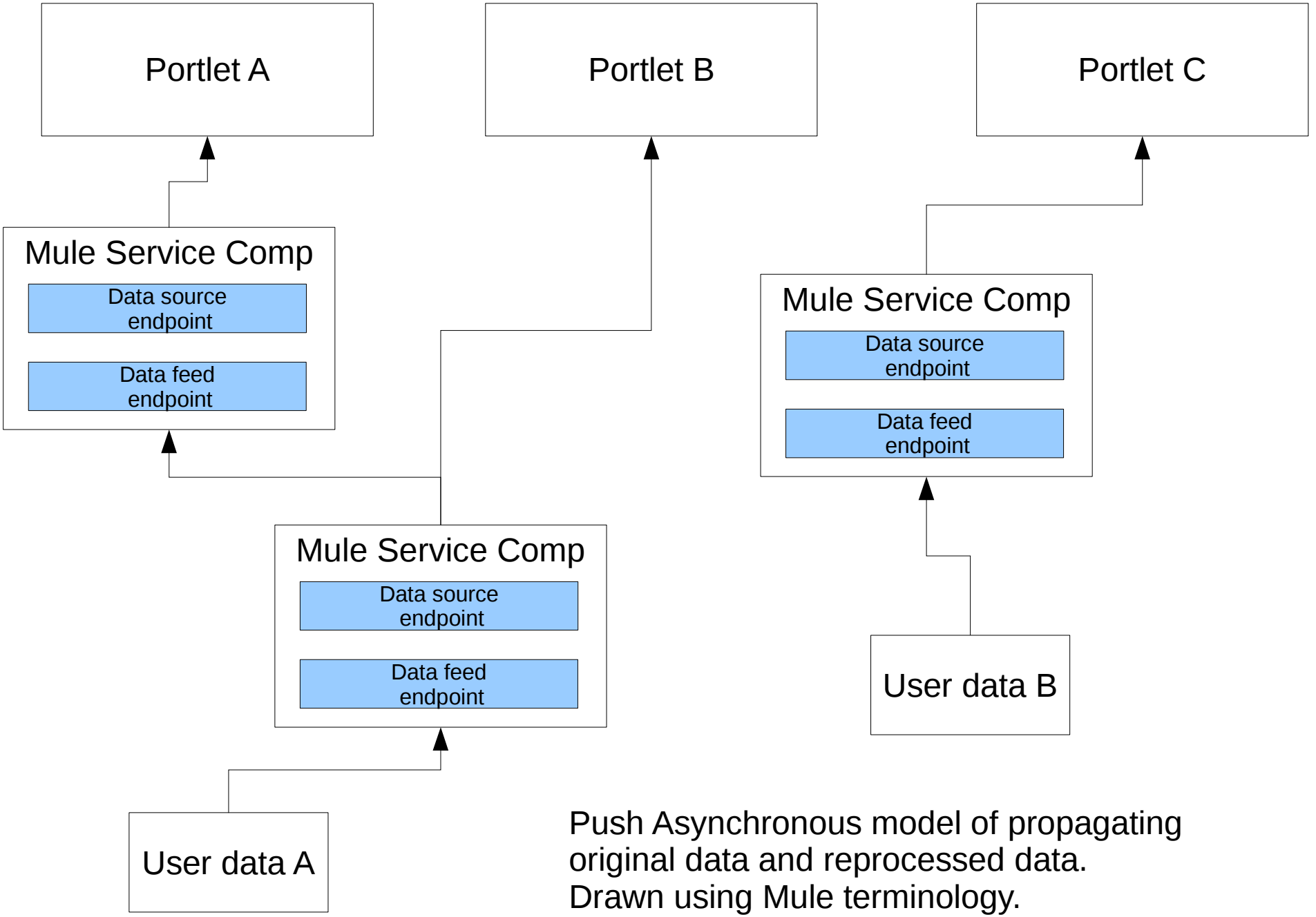
User perspective

- Use DS admin tool to define:
  - Schema, ds name, inbound transformation,transformation parameters
- Use Workflow admin tool to define
  - Connections between data source/feed, outbound transformations,transformation parameters

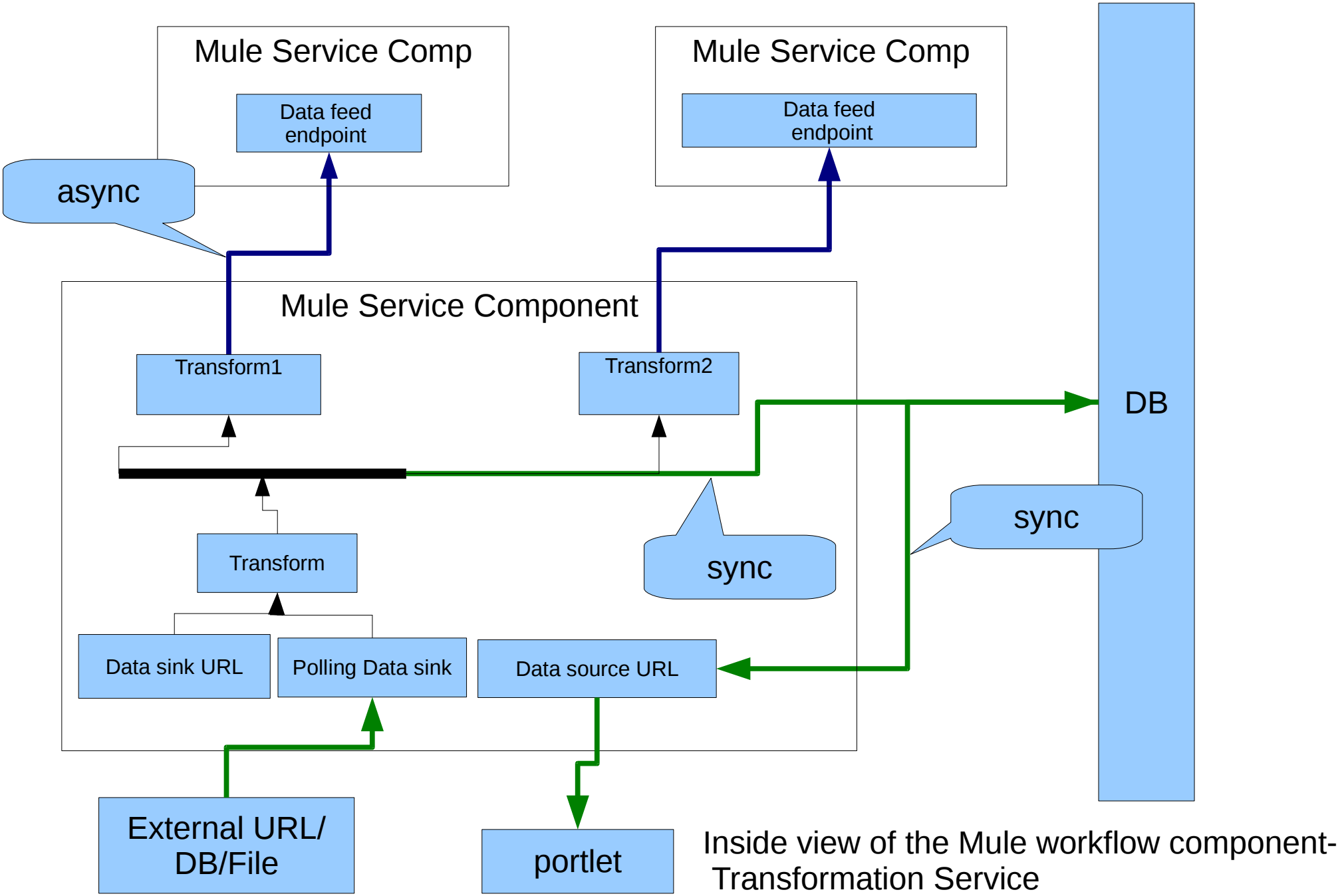
System perspective



- Use DS Api to define logical rep. of data sources and their workflow connections
- Use offline configurator to instantiate Mule configuration corresponding to desired workflow
- Restart Mule



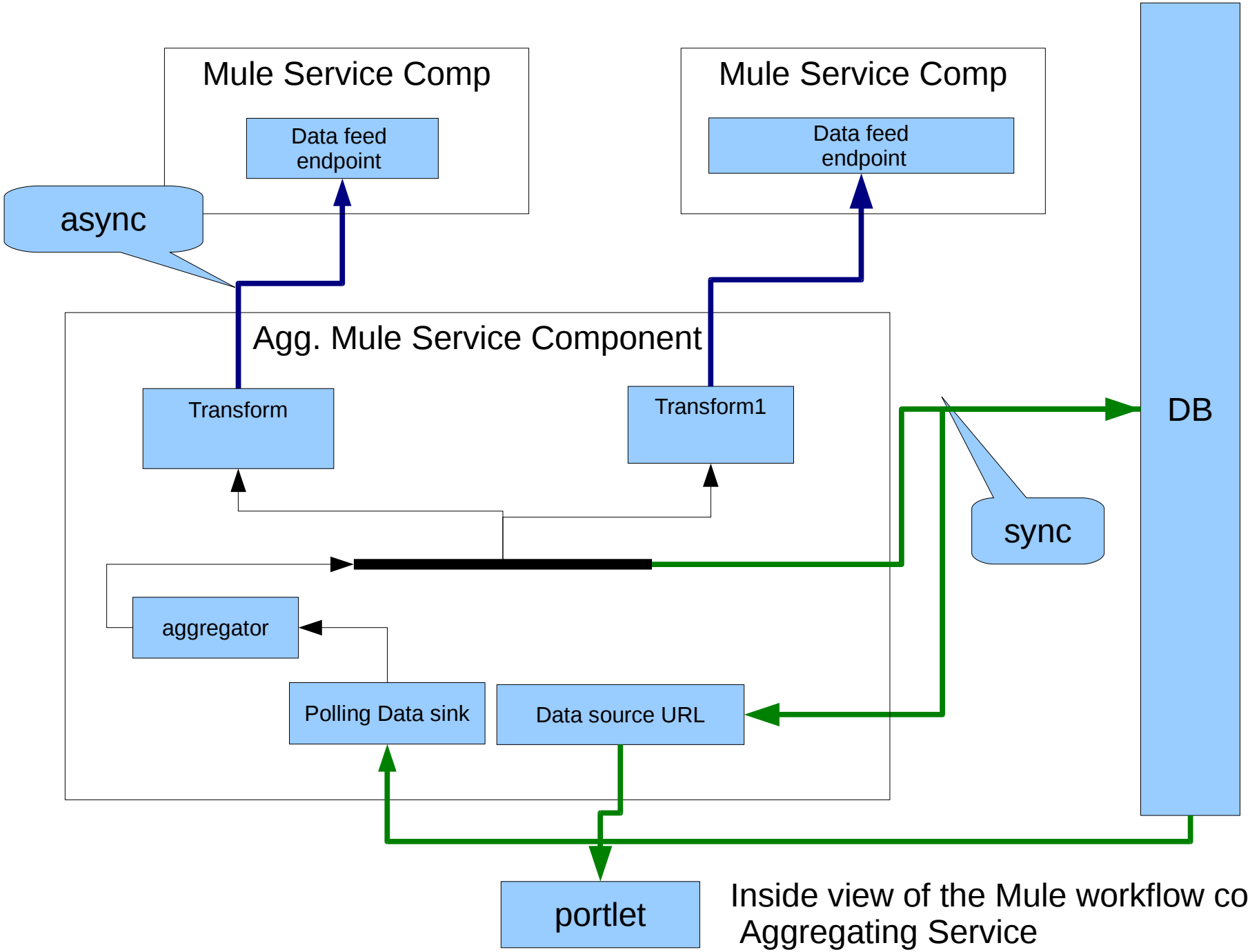
Push Asynchronous model of propagating original data and reprocessed data.
Drawn using Mule terminology.



Inside view of the Mule workflow component- Transformation Service

Mule mapping. Transformation service.

- Service component encapsulates
  - Transformations related to particular data schema
  - Workflow connections
  - Synchronism model
  - Access protocol
- Data is accepted through polling or it can be posted directly via “data” sink.
- Data is published via “feeds” after applying transformation, particular to that feed.
- Transformation can ANYTHING
  - XSLT, java script , java code, java scripting code.



Inside view of the Mule workflow component- Aggregating Service

Mule mapping.Aggregating component.

- “Aggregating” component is “special” because it does not accept user data directly
- Instead data is pulled from XML DB (or other DB) , processed(aggregated) and then pushed back to DB AND other service components.